Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition

Petkov, Nicolai; Subramanian, Easwar

Published in:
Biological Cybernetics

DOI:
10.1007/s00422-007-0182-0

Publication date:
2007

Biol Cybern (2007) 97:423–439
DOI 10.1007/s00422-007-0182-0

ORIGINAL PAPER

Biological Cybernetics

Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition

Nicolai Petkov · Easwar Subramanian

Received: 28 May 2007 / Accepted: 21 September 2007 / Published online: 25 October 2007

Abstract We study the orientation and speed tuning properties of spatiotemporal three-dimensional (3D) Gabor and motion energy filters as models of time-dependent receptive fields of simple and complex cells in the primary visual cortex (V1). We augment the motion energy operator with surround suppression to model the inhibitory effect of stimuli outside the classical receptive field. We show that spatiotemporal integration and surround suppression lead to substantial noise reduction. We propose an effective and straightforward motion detection computation that uses the population code of a set of motion energy filters tuned to different velocities. We also show that surround inhibition leads to suppression of texture and thus improves the visibility of object contours and facilitates figure/ground segregation and the detection and recognition of objects.

Keywords V1 · Gabor filter · Receptive field · Simple cell · Complex cell · Motion energy · Spatiotemporal integration · Surround suppression · Contour detection · Texture suppression · Noise reduction · Motion detection · Figure/ground segregation

1 Introduction

The visual system of man and animals has been the subject of intense research for several decades. An important finding in the neurophysiology of the visual system of cats and monkeys, made at the beginning of the 1960s, was that the majority of neurons in the primary visual cortex (V1) respond to a line or an edge of a certain orientation in a given position of the visual field (Hubel and Wiesel 1962; Hubel and Wiesel 1968). Primarily, two types of orientation selective neurons were identified, one that was sensitive to the contrast polarity of bars and edges, called simple cells, and another that was not, called complex cells. Computational models were developed aiming at simulating the function of these neurons for understanding and predicting their responses to more complex visual stimuli. The spatial summation properties of simple cells were modeled by linear filters followed by half-wave rectification (Movshon et al. 1978b; Andrews and Pollen 1979; Glezer et al. 1980; Kulikowski and Bishop 1981) and Gabor functions proved to be particularly well suited for this purpose (Marcelja 1980; Daugman 1985; Jones and Palmer 1987). Complex cells needed more intricate modeling, which included linear filtering, half-wave rectification and subsequent local spatial summation, or quadrature pair summation of linear filter responses (Movshon et al. 1978a; Spitzer and Hochstein 1985; Morrone and Burr 1988; Petkov and Kruizinga 1997; Kruizinga and Petkov 1999; Grigorescu et al. 2002, Grigorescu et al. 2003). These computational models contributed to understanding of the functions of simple and complex cells and gave the basis for biologically motivated edge detection algorithms in image processing and computer vision (see Fig. 1).

However, most of these studies were based on the spatial properties of the receptive field (RF) organization. Later, sophisticated RF mapping techniques revealed that the RFs of cortical cells change in time and hence they must be considered as spatiotemporal entities. Indeed, the RF profiles of many simple cells are inseparable functions of space and time, and their specific structure of alternating elongated excitatory and inhibitory regions that are tilted with respect

N. Petkov (✉) · E. Subramanian
Institute of Mathematics and Computing Science,
University of Groningen, P. O. Box 407,
9700 AK Groningen, The Netherlands
e-mail: petkov@cs.rug.nl

E. Subramanian
e-mail: easwar@cs.rug.nl

Fig. 1 Edge and contour detection in the spatial domain. The input image (*left*) is processed by a Gabor energy operator that is motivated by the function of complex cells. The binarized output of that operator is shown in the middle image. The operator essentially acts as an edge detector and does not distinguish the edges that belong to the contour of the animal from those of the background texture. The *right image* shows the binarized output of a Gabor energy operator that is augmented with surround suppression (Petkov and Westenberg 2003; Grigorescu et al. 2003). The contours of the animal are better visible in this image due to the removal of the texture edges by means of surround inhibition

to the time axis underlie the speed and direction selectivity of these cells (DeAngelis et al. 1993a, b, 1995). Therefore, these V1 cells are essentially spatiotemporal filters and they combine information over space and time.

One of the apparent advantages of a spatiotemporal filter over a spatial filter is that the former can be used for motion analysis. A purely spatial filter cannot be used for this purpose because it considers information only at a single time instant, while motion is a spatiotemporal concept implying changes over time. Since a stimulus in a given position will evoke responses in a number of cells whose receptive fields include that position, it is interesting to know how motion is coded in the group of these responses. One purpose of this study is to take a closer look at population coding by spatiotemporal filters and to see whether it allows the extraction of motion attributes such as the presence or absence of motion at a given position.

As to the processing of image sequences for edge detection, one can apply a spatial filter on a frame-by-frame basis or a spatiotemporal filter that uses information within and across frames. Another purpose of this work is to closely examine the benefits of using spatiotemporal filters instead of purely spatial filters to process image sequences.

Furthermore, neurophysiological studies have also showed that, once a cell is activated by a stimulus in its classical receptive field (CRF), another, simultaneously presented stimulus outside that field can have an effect on the cell response (Blakemore and Tobin 1972; Knierim and van Essen 1992; Nothdurft et al. 1999; Jones et al. 2001). This, mostly inhibitive, effect is known as *nonclassical receptive field inhibition* or *surround suppression*. With respect to the spatial properties of simple and complex cells in V1, surround inhibition[1] is an useful mechanism for contour detection by suppression of texture (Petkov and Westenberg 2003; Grigorescu et al. 2003) and has been applied to other features as well (Rodrigues and du Buf 2005a, b; Rodrigues and du Buf 2006). Its application to contour detection is illustrated in Fig. 1, where the input image shown on the left is processed by a Gabor energy operator that is motivated by the function of complex cells. The binarized output of that operator is shown in the middle image. The operator essentially acts as an edge detector and does not distinguish the edges that belong to the contour of the animal from those of the background texture. The right image shows the binarized output of a Gabor energy operator that is augmented with surround suppression. The contours of the animal are more visible in this image due to the removal of the texture edges by means of surround inhibition. A similar mechanism has been observed in the spatiotemporal domain (Allman et al. 1985) and it is known to have several functional implications to motion processing (Born and Bradley 2005). A further aim of the current work is to explore some functional aspects of surround suppression in motion processing using a computational model.

Surround interactions are observed in different cortical regions such as V1 Jones et al. 2001, middle temporal (MT/V5) (Allman et al. 1985; Raiguel et al. 1995) and lateral medial superior temporal (MST) (Eifuku and Wurtz 1998), which are areas involved in processing motion information. Also, it is known that the RFs of about one half of the cells in MT have antagonistic surrounds (Allman et al. 1985; Tanaka et al. 1986; Raiguel et al.1995; Bradley and Anderson 1998; Born 2000; DeAngelis and Uka 2003; Born and Bradley 2005). The response of such a neuron is suppressed when moving stimuli are presented in the region surrounding its CRF. The suppression is maximal when the surround stimuli move in the same direction and at the same disparity as the preferred center stimulus (Allman et al. 1985; Born and Bradley 2005; Bradley and Anderson 1998; Raiguel et al. 1995). In addition, neurons with facilitative surround struc-

[1] Throughout this text we use the words inhibition and suppression as synonyms.

tures have also been found (Allman et al. 1985; Born and Tootell 1992; Raiguel et al. 1995). Such neurons show an increased response when motion is presented to their surround and are found in locations that are anatomically different from the ones that have antagonistic surrounds (Born and Tootell 1992). Moreover, surround mechanisms differ for low- and high-contrast stimuli (Tadin et al. 2003; Pack et al. 2005; Paffen et al. 2005): facilitation happens at low contrast and suppression occurs at high contrast.

An important utility of surround mechanisms in the spatiotemporal domain is to detect motion discontinuities or motion boundaries (Nakayama and Loomis 1974). The functional role also depends on the spatial organization of the surround. Neurons with a symmetric surround are hypothesized to play a role in figure/ground segregation (Born and Bradley 2005). Asymmetric surround structures are thought to aid in determining surface tilt (or slant) and curvature (Koenderink and van Doorn 1992; Buracas and Albright 1996). The surround mechanisms are also thought to be involved in motion segmentation and shape-from-motion processing Gautama and van Hulle (2001). In the current work, we closely examine the role of center–surround interactions in the context of texture suppression and contour enhancement. To this end, we first describe a computational model to process visual motion and augment it with a surround suppression term to qualitatively reproduce the center–surround behavior of motion-sensitive neurons.

The paper is organized as follows. In Sect. 2 we begin with an outline of the computational models of the CRFs of motion-sensitive V1 cells. We consider spatiotemporal motion energy filters and examine their direction and speed tuning properties. Then we augment these filters with a surround suppression computation. In Sect. 3 we analyze the utility of the proposed operators for noise reduction, motion detection, texture suppression, and improving contour visibility. Section 4 contains a discussion on various aspects of the model. Finally, in Sect. 5, we present our conclusions. The Appendices contain the mathematical details of the proposed operators.

2 Computational models

2.1 Spatiotemporal Gabor filters

In a seminal work, (Adelson and Bergen 1985) suggested that a two-dimensional (2D) spatial pattern moving at a given velocity corresponds to a three-dimensional (3D) spatiotemporal pattern of a given orientation which can be detected with an appropriately oriented 3D spatiotemporal filter, such as a 3D Gabor filter. To this end, we model the spatiotemporal receptive field profiles of simple cells as a family of 3D Gabor functions denoted by $g_{v,\theta,\varphi}(x, y, t)$ where the parameter v is the preferred speed, the angle parameter θ determines the preferred direction of motion and the preferred spatial orientation of the filter, and φ is a parameter that determines the spatial symmetry of the function. Essentially, $g_{v,\theta,\varphi}(x, y, t)$ is a product of a Gaussian envelope function that restricts $g_{v,\theta,\varphi}(x, y, t)$ in the spatial domain, a cosine wave traveling with a phase speed v in direction θ, another Gaussian function that depends only on the time t and determines the temporal decay of $g_{v,\theta,\varphi}(x, y, t)$, and a step function of t that ensures that the filter based on $g_{v,\theta,\varphi}(x, y, t)$ is causal and thereby considers inputs only from the past. The mathematical details are provided in the Appendix A.

In Fig. 2, the space–time profiles of $g_{v,\theta,\varphi}(x, y, t)$ are rendered for a stationary Gaussian envelope and an envelope that moves together with the cosine wave. Also shown are x–t plots of spatiotemporal RF profiles computed with $y = 0$ in $g_{v,\theta,\varphi}(x, y, t)$. These plots are qualitatively similar to the experimentally determined ones by DeAngelis and co-workers (DeAngelis et al. 1993a, b, 1995). The tilt of the excitatory and inhibitory subregions in the space–time domain is the origin of the selectivity for moving stimuli that leave similar tilted traces in space–time. Our main motivation for introducing two types of spatiotemporal RF profiles as shown in Fig. 2 (i.e., stationary and moving envelope) is to explore if there are any significant qualitative differences in the computational properties of one model over the other, to examine the plausibility of suggestions previously made in this context (DeAngelis et al.1993a, b, 1995; van Hateren and Ruderman 1998) and to check if this is an issue of importance.

We compute the spatial period or wavelength λ of the cosine wave using the following function of v: $\lambda = \lambda_0\sqrt{1+v^2}$, where λ_0 is the spatiotemporal period of the filter. The above relation implies that filters that prefer higher speeds have bigger receptive fields. In Fig. 3, x–t plots are rendered for cells preferring rightward motion ($\theta = 0$) at four different speeds $v \in \{0, 1, 2, 4\}$ for the moving and the stationary envelope cases. Observe that as the speed increases the subregions are tilted more towards the axis of movement (here the x axis). The larger the preferred speed v, the larger is the spatial period of the wave along that axis.

The response $r_{v,\theta,\varphi}(x, y, t)$ of a linear filter with a RF function $g_{v,\theta,\varphi}(x, y, t)$ to a luminance distribution $I(x, y, t)$ is computed by convolution:

$$r_{v,\theta,\varphi}(x, y, t) = I(x, y, t) * g_{v,\theta,\varphi}(x, y, t). \quad (1)$$

The response of a model simple cell with a RF centered on (x, y) at time t is computed from the linear response $r_{v,\theta,\varphi}(x, y, t)$ using half-wave rectification:

$$s_{v,\theta,\varphi}(x, y, t) = |r_{v,\theta,\varphi}(x, y, t)|^+ \quad (2)$$

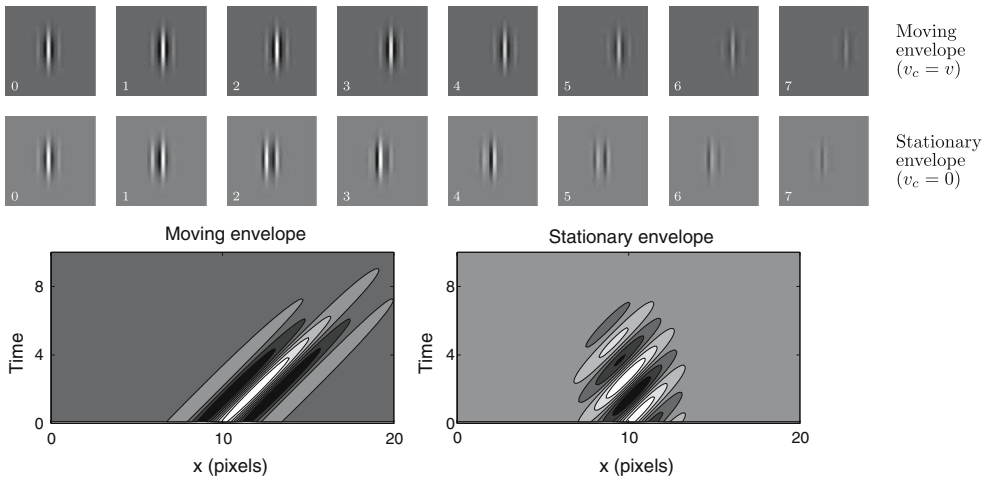

Fig. 2 Spatiotemporal behavior of $g_{v,\theta,\varphi}(x, y, t)$ for $v = 1$ (in pixels per frame), $\theta = 0$, and $\varphi = 0$. Two types of spatiotemporal RF profiles are shown. The *top row* contains the profile of a filter where the spatial Gaussian envelope moves at speed v_c, which is equal to the speed v of the traveling cosine wave (i.e., $v = v_c$). The *second row* shows the profile of a filter with a stationary spatial Gaussian envelope (i.e., $v_c = 0$). In each frame, the x–y profile at a particular time instant is shown in the two rows with elongated light and dark regions representing excitatory and inhibitory lobes of the filter, respectively. The preferred direction of movement ($\theta = 0$) is perpendicular to these regions. Below these rows are the x–t plots of the respective RF profiles. One can observe that the excitatory and inhibitory subregions are tilted in the space–time domain toward the direction of movement (here the x axis). A *light bar* stimulus oriented parallel to the y axis and traveling along the x axis will leave a trace in the x–t domain that is similar to the excitatory lobes of the shown spatiotemporal receptive fields and will elicit strong responses in the corresponding model cells. The purpose of discussing two types of spatiotemporal RF profiles is to explore if there are any significant qualitative differences in the computational properties of one model over the other

where $|.|^+$ is defined as follows:

$$|z|^+ = \begin{cases} z & \text{if } z \geq 0 \\ 0 & \text{if } z < 0. \end{cases} \tag{3}$$

A simple cell is phase sensitive in the sense that its response to a moving pattern depends on the stimulus contrast polarity and exact position within the receptive field. This property is reproduced by the computational model according to equations (1)–(2). A phase insensitive response can be obtained by quadrature pair summation of the responses of two filters with a phase difference of $\pi/2$ as follows:

$$E_{v,\theta}(x, y, t) = \sqrt{r^2_{v,\theta,0}(x, y, t) + r^2_{v,\theta,\pi/2}(x, y, t)}. \tag{4}$$

This quantity, called motion energy (Adelson and Bergen 1985), is phase insensitive and can be used as a model of the response of a complex cell.

2.2 Direction and speed tuning properties

In the following part, we briefly examine the direction and speed tuning properties of the motion energy filter described above. For direction tuning, we consider bars moving at the same speed but in different directions θ_s (see Fig. 4). For each stimulus, we compute the response of filters which have preference for the same speed but are tuned to different directions. The maximum response is obtained when the preferred direction of the filter (θ) matches the direction of movement of the bar (θ_s), as seen in the diagonal entries.

The speed tuning properties are studied by considering the responses of motion energy filters to edges drifting rightward at different speeds (see Fig. 5). For this experiment, we choose filters that have a preference for the same direction of motion ($\theta = 0$) but differ in their preference for speed. Maximum response is obtained when the preferred speed of the filter (v) matches the speed of the edge (v_s), as seen in the diagonal entries.

The direction and speed tuning properties can also be depicted as in Fig. 6. In Fig. 6a we show a plot of the maximum response of each filter, at a particular frame, to a vertical bar moving rightward ($\theta_s = 0$) at a speed of one pixel per frame ($v_s = 1$). The response reaches its maximum when the direction of movement of the stimulus matches the preferred direction of motion of the filter. A similar plot is also shown for speed tuning (Fig. 6b), where the stimulus is an edge drifting rightward ($\theta_s = 0$) at a speed of two pixels per frame ($v_s = 2$). The response reaches its peak when the phase speed of the traveling cosine wave is equal to the speed of the stimulus. For this reason, the phase speed v of the traveling cosine wave can be considered as the preferred speed of motion of the filter. From Fig. 6, one can observe that a filter with a moving envelope (solid line) is more selective for direction and speed than a filter with a stationary envelope (dashed line). For this reason, in all subsequent experiments we choose to work with filters with a moving envelope. As we

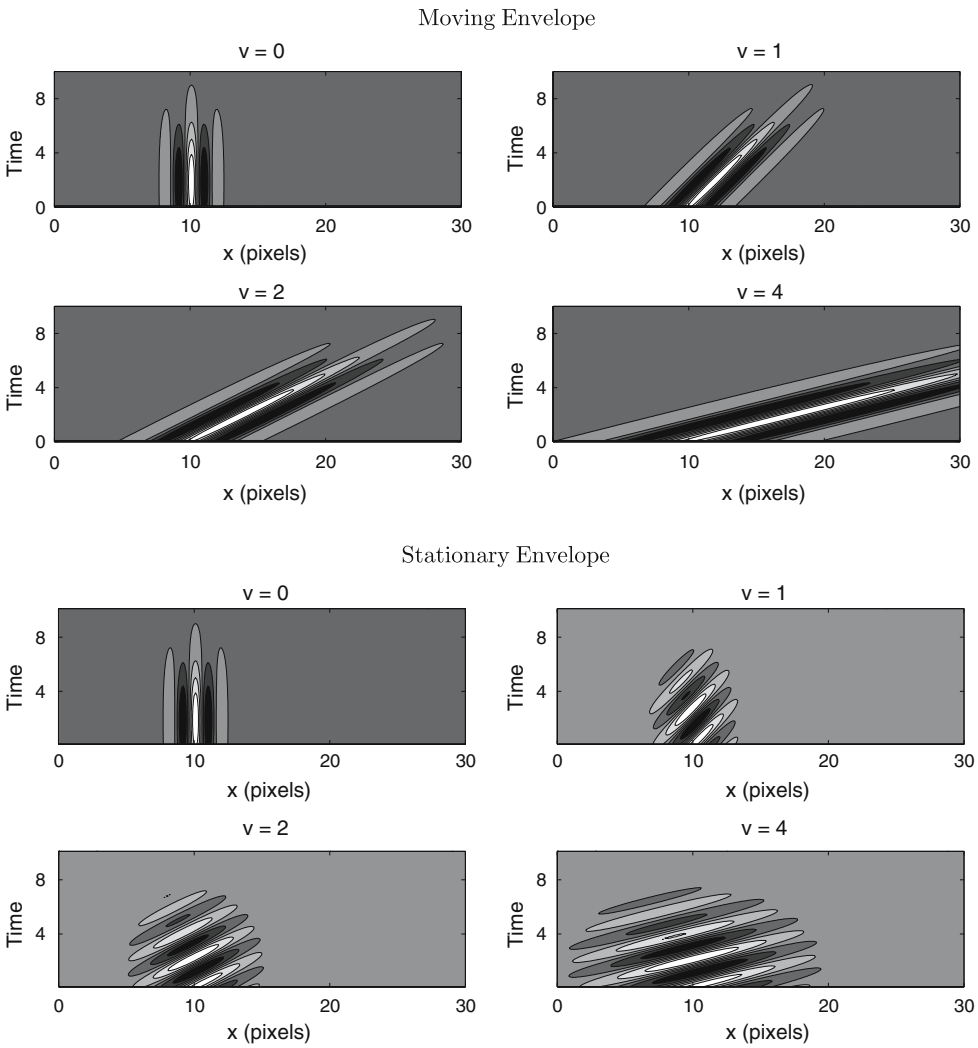

Fig. 3 x–t plots for cells preferring rightward motion ($\theta = 0$) at four different speeds $v \in \{0, 1, 2, 4\}$ (in pixels per frame) for the moving envelope (*upper block*) and the stationary envelope (*lower block*) cases. Observe that as the speed increases the subregions are tilted more towards the axis of movement (here the x axis) and the spatial period λ of the wave along that axis increases

see from Fig. 6b the speed of the envelope has no influence on the preferred speed of the filter.

2.3 Surround suppression model

In this section we propose a surround inhibition operator that takes into account the influence of the surround at each spatial location and time instant. It is a straightforward generalization of a model that was used in the case of a purely spatial filter (Petkov and Westenberg 2003; Grigorescu et al. 2003, 2004). The classical receptive field (CRF) of a model simple cell is defined as the area in which the (moving) Gaussian envelope of the corresponding 3D Gabor function $g_{v,\theta,\varphi}(x, y, t)$ is substantial. It contains all points within a certain Mahalanobis distance (Mahalanobis 1936) from the center of that envelope. We define the surround suppression weighting function $w_{v,\theta}(x, y, t)$ to be zero inside the CRF and positive outside it and to decay with the distance to the CRF (see Fig. 7). In practice, we take as a surround weighting function the half-wave-rectified difference of two concentric Gaussian envelopes, of which one is identical to what was used in the CRF function $g_{v,\theta,\varphi}(x, y, t)$, while the other has a spatial extent that is several times larger. Furthermore, the surround weighting function decays with time in the same way as the CRF function $g_{v,\theta,\varphi}(x, y, t)$. The mathematical details are provided in the Appendix B. In Fig. 8, we render the x–t plot of $w_{v,\theta}(x, y, t)$ for $y = 0$.

For each point in the (x, y, t) space, we compute an inhibition term $S_{v,\theta}(x, y, t)$ by weighted summation of the motion energy $E_{v,\theta}(x, y, t)$ in the surroundings of that point using the surround weighting function $w_{v,\theta}(x, y, t)$. In practice, the inhibition term is computed by convolution:

$$S_{v,\theta}(x, y, t) = E_{v,\theta}(x, y, t) * w_{v,\theta}(x, y, t). \quad (5)$$

The larger and denser the motion energy $E_{v,\theta}(x, y, t)$ in the surroundings of a point (x, y, t), the larger the suppression term $S_{v,\theta}(x, y, t)$ is at that point. We next use this inhibition

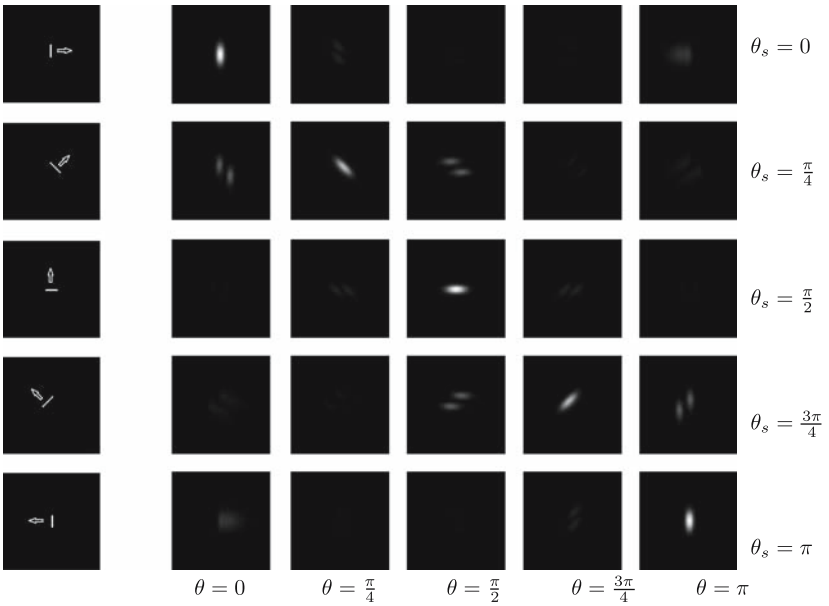

Fig. 4 Responses of motion energy filters to moving bars. Snapshots of the stimuli, *bars* moving at a speed $v_s = 1$ (in pixels per frame), in various directions θ_s, are shown in the leftmost column. Each of the other columns show snapshots of the response $E_{v,\theta}(x, y, t)$ of a filter with a given preferred orientation θ specified at the bottom of the column. All filters have preference for the same speed (i.e., $v = 1$) but differ in their preference for direction of motion. A stimulus moving in a given direction θ_s elicits the strongest response in a filter preferring the same direction of motion $\theta = \theta_s$ (see diagonal entries). Since the responses of filters with stationary ($v_c = 0$) and moving envelopes ($v_c = v$) are visually similar, we choose to show only the responses of the moving envelope case

term to define and compute a surround suppressed motion energy $\tilde{E}_{v,\theta}(x, y, t)$ as follows:

$$\tilde{E}_{v,\theta}(x, y, t) = |(E_{v,\theta}(x, y, t) - \alpha S_{v,\theta}(x, y, t))|^{+}, \qquad (6)$$

where the factor α controls the strength with which surround suppression is taken into account. The proposed inhibition scheme is a subtractive linear mechanism followed by a nonlinear half-wave rectification. Note that in each point, the motion energy response for a given preferred speed v and orientation θ is suppressed only by responses for the same preferred speed and orientation in the surround of that point. Since the motion energy filters are broadly tuned to orientation and speed, stimuli with a broad range of orientations and speeds will have an inhibitory effect. However, the suppression will be strongest when the stimuli in the surroundings of a point have the same direction and speed of movement as the stimulus in the concerned point. In reality, a neuron tuned to a certain velocity and orientation may be inhibited by other neurons tuned to nearby velocities and orientations. As a result, the suppression will be minimal when the surround stimuli move in opposite direction as compared to the stimulus in the center. This aspect of our model corresponds to neurophysiological findings concluding that surround stimuli with the same direction and speed as the optimal CRF stimulus have a larger suppressive effect on the response of a motion selective neuron than stimuli of other directions and speed of motion (Allman et al. 1985; Raiguel et al. 1995; Bradley and Anderson 1998). Our model (of V1 cells) also bears a certain resemblance to MT cells, for which the efficacy of center–surround interactions is increased by opposite motion directions.

The effect of surround suppression is illustrated in Fig. 9. The stimulus, which is shown in the leftmost image, consists of a bar grating and one isolated bar. While the grating is moving rightward ($\theta_s = 0$), the isolated bar is moving leftward ($\theta_s = \pi$). Both are moving at a constant speed of one pixel per frame ($v_s = 1$). Subsequent entries show snapshots of the superimposed responses of the motion energy filter, the inhibition term, and the motion energy operator augmented with surround suppression, respectively. While the response of the motion energy filter is similar to that of the isolated bar and to the bars that form the grating, the surround suppressed motion energy responds only to the bar that is not surrounded by other similar stimuli. A similar result is obtained when the bar and the grating move in the same direction. In this way, surround mechanisms help separate objects (the isolated bar) from their backgrounds (grating). As we shall show later, this property of the surround suppressed motion energy operator leads to improved visibility of object contours and region boundaries and thus makes it more effective for object recognition. The surround suppressed motion energy operator inherits the properties of the motion energy operator with respect to speed and orientation tuning.

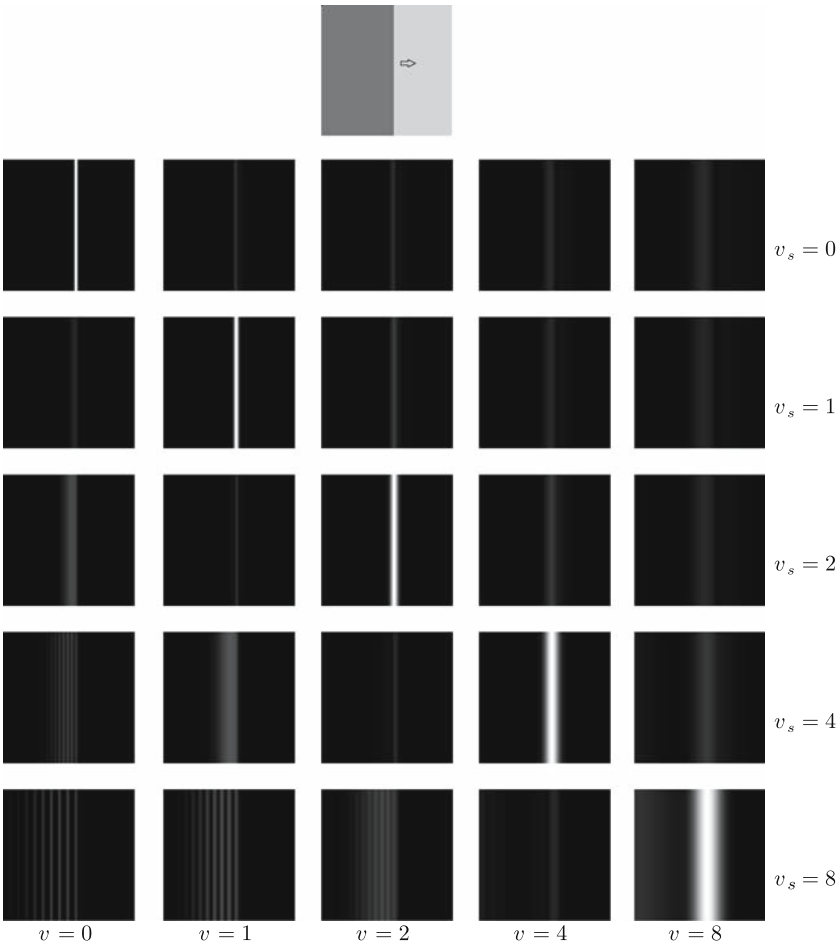

Fig. 5 Responses of motion energy filters to edges drifting at different speeds. A snapshot of the stimulus, an edge drifting rightward ($\theta_s = 0$) at a given speed, is shown in the top row. Each subsequent entry is a snapshot of the response $E_{v,\theta}(x, y, t)$ of a filter with a speed v (in pixels per frame) specified at the end of each column to an edge that is drifting at a particular speed v_s indicated at the end of each row. All filters have preference for the same direction of motion ($\theta = 0$) but differ in their preference for speed. An edge drifting at a given speed v_s elicits strongest response in the filter with the same preferred speed $v = v_s$ (see diagonal entries). Since the responses of filters with stationary ($v_c = 0$) and moving envelopes ($v_c = v$) are visually similar, we choose to show only the responses of the moving envelope case

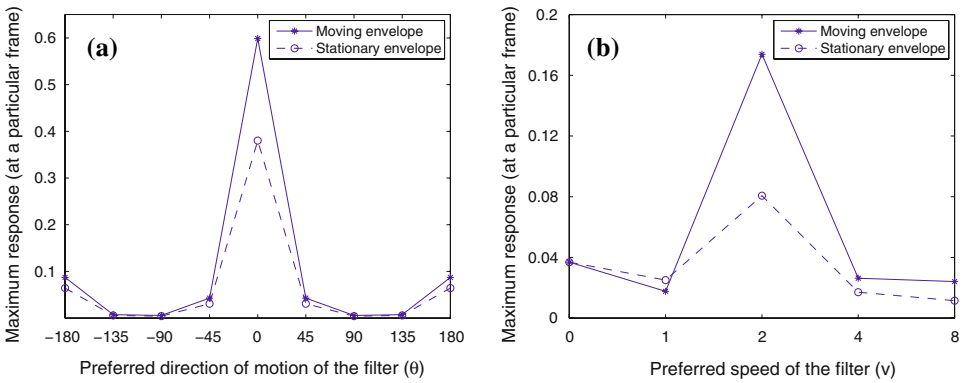

Fig. 6 **a** Direction tuning properties of motion energy filters: the maximum response of each filter, at a particular frame, to a moving bar ($\theta_s = 0$, $v_s = 1$) as a function of the difference between the direction of bar movement and the preferred direction of the filter. The peak response is obtained for the filter whose preferred direction matches the direction of the stimulus. **b** Speed tuning: the maximum response of each filter, at a particular frame, to a drifting edge ($\theta_s = 0$; $v_s = 2$) as a function of the preferred speed of the filter. The response reaches the peak when the speed of the stimulus matches the preferred speed of the filter. The responses were computed for filters with a moving (*solid line*) and a stationary (*dashed line*) envelope

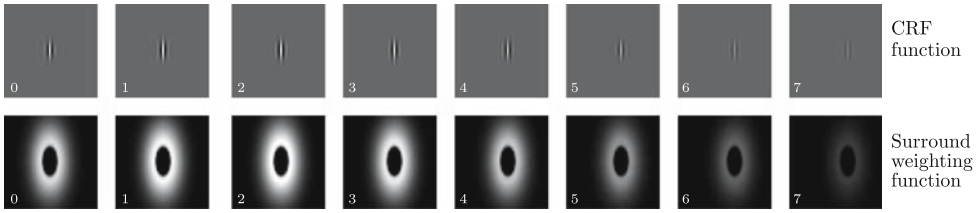

Fig. 7 Spatiotemporal behavior of the CRF function $g_{v,\theta,\varphi}(x, y, t)$ and the corresponding surround weighting function $w_{v,\theta}(x, y, t)$ with $v = 1$ (in pixels per frame) and $\theta = 0$ for the moving envelope case. The first row contains the profile of $g_{v,\theta,\varphi}(x, y, t)$ with $\varphi = 0$; the second row contains the profile of $w_{v,\theta}(x, y, t)$

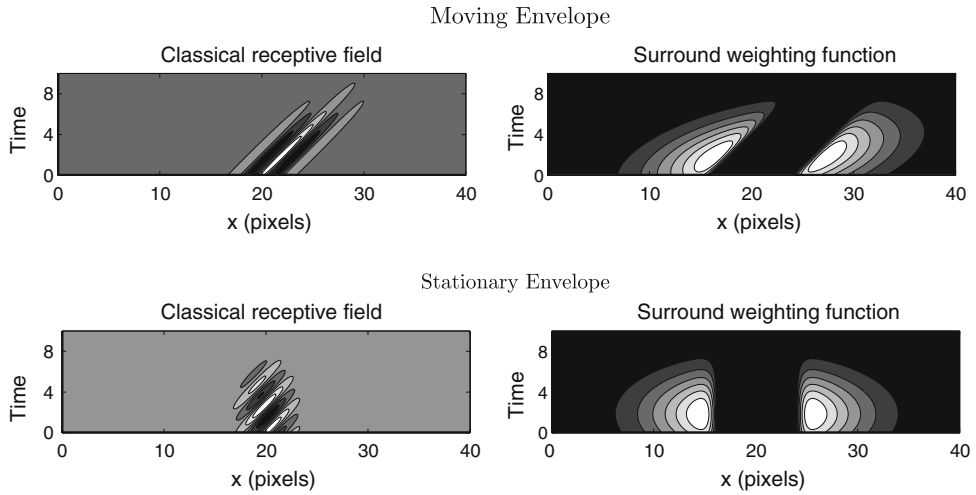

Fig. 8 x–t plots for the classical receptive field function $g_{v,\theta,\varphi}(x, y, t)$ and the corresponding surround weighting function $w_{v,\theta}(x, y, t)$ for the moving (*upper block*) and the stationary (*lower block*) envelope cases for $\theta = 0$, $v = 1$, and $\varphi = 0$

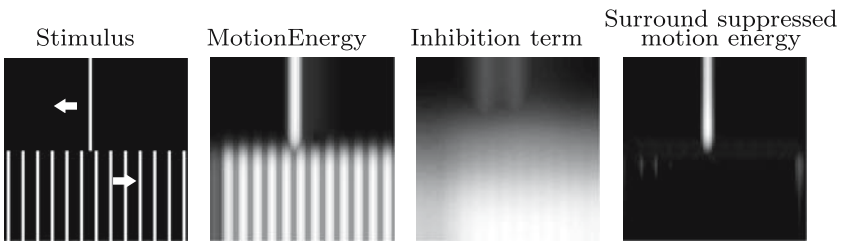

Fig. 9 Effect of the surround suppression operator: the stimulus whose snapshot is shown in the leftmost image comprises a grating and one isolated bar. The grating is moving rightwards ($\theta_s = 0$) at a constant speed ($v_s = 1$) and the isolated bar is moving leftwards ($\theta_s = \pi$) at the same speed ($v_s = 1$). Subsequent entries show the snapshots of superposed responses of motion energy filter for $\theta = 0$ and $\theta = \pi$, the inhibition term and surround suppressed motion energy ($\alpha = 2$), respectively. While the response of the motion energy filter is alike to the isolated bar and to the bars that form the grating, the surround suppressed motion energy operator responds only to the bar that is not surrounded by other stimuli. In this way, surround mechanisms help separate objects (here an isolated bar) from their backgrounds (texture represented here by a grating)

3 Benefits of spatiotemporal integration and surround suppression

3.1 Noise suppression

Spatiotemporal integration and surround suppression enhances robustness to noise. In Fig. 10 we illustrate this idea using a drifting bar stimulus with added random Gaussian noise. In addition the bar is broken in the 12th, 22nd, and 34th time units. Subsequent rows contain the responses obtained from a purely spatial Gabor energy (Grigorescu et al. 2003), spatiotemporal motion energy, and surround suppressed motion energy filters. The response of the spatial Gabor energy filter, shown in the second row, is obtained by taking into account only the input image at the corresponding current time. One can observe that, unlike the Gabor energy filter, spatiotemporal filters, by integrating inputs over time, significantly reduce the noise and restore the integrity of the

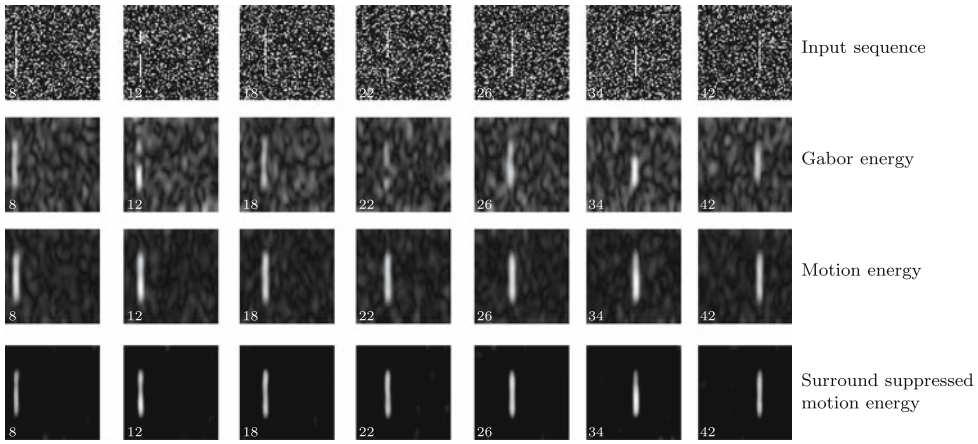

Fig. 10 Noise reduction. The *first row* shows snapshots of a bar moving rightward with a speed of one pixel per frame. Uncorrelated random Gaussian noise is added to each frame and in addition the bar is broken in the 12th, 22nd, and 34th frames. Subsequent rows contain the responses obtained from a spatial Gabor energy (*second row*), a motion energy (*third row*), and a surround suppressed motion energy ($\alpha = 2$, *fourth row*) filters. By integrating inputs over time, spatiotemporal operators significantly reduce the noise and restore the integrity of the bar. Further noise reduction is due to the surround suppression mechanism

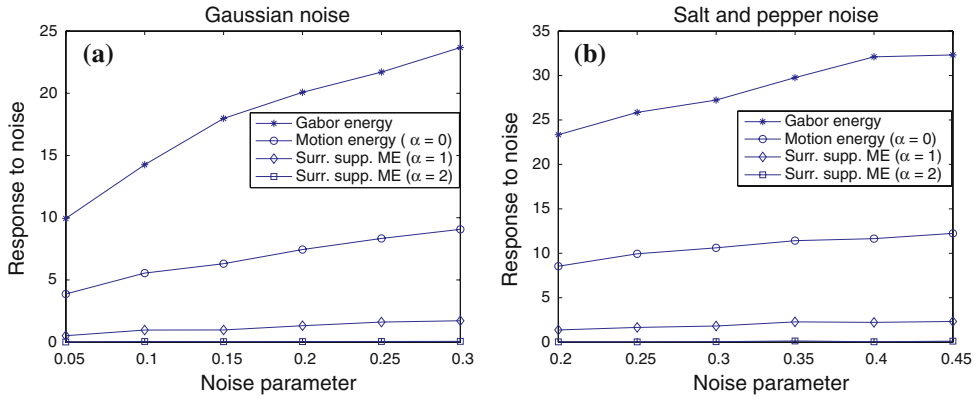

Fig. 11 Noise reduction in spatiotemporal filters for two different noise types of varied strengths. **a** Gaussian noise with mean zero and variance as shown in the x axis. **b** Salt-and-pepper noise with noise density in the x axis. Uncorrelated random noise is added to each frame of the input sequence. The response of an operator to noise was calculated as the sum of the values obtained in a region that does not contain the stimulus, averaged over all frames. For surround suppression with $\alpha \geq 2$ the noise is practically eliminated

bar. This is due to the fact that, while noise is uncorrelated from frame to frame, the signal shifts at a constant speed and, provided that an appropriate motion energy filter is used, the current and past frames combine information in a coordinated way to form the current output frame. The improved response from the surround suppressed motion energy operator is due to the inhibition mechanism where the noise in one position is inhibited by noise in neighboring regions. We carried out a quantitative study using two different noise types with varying noise strengths and different values of the surround suppression parameter α. We calculated the response of an operator to noise as the sum of the values obtained in a region that does not contain the stimulus. The results shown in Fig. 11 indicate that in all cases the surround mechanisms are very effective in noise reduction. For $\alpha \geq 2$, the noise is practically eliminated.

3.2 Motion detection

A purely spatial filter computes the output at a given time using only the input at that time. Hence, it cannot be used for motion analysis because in image sequences motion manifests itself in changes in space and time. In the case of a spatiotemporal filter, inputs from the present and the past are used to compute the response at the current moment and hence such a filter can be used for analyzing and detecting motion.

Motion detection is a basic task that is performed by the visual system and it is therefore interesting to examine more closely the relation of the responses of the considered spatiotemporal filters to that task. As illustrated in Figs. 5 and 6b, the considered filters are not sharply tuned to one single speed: a filter that prefers stationary stimuli will also respond

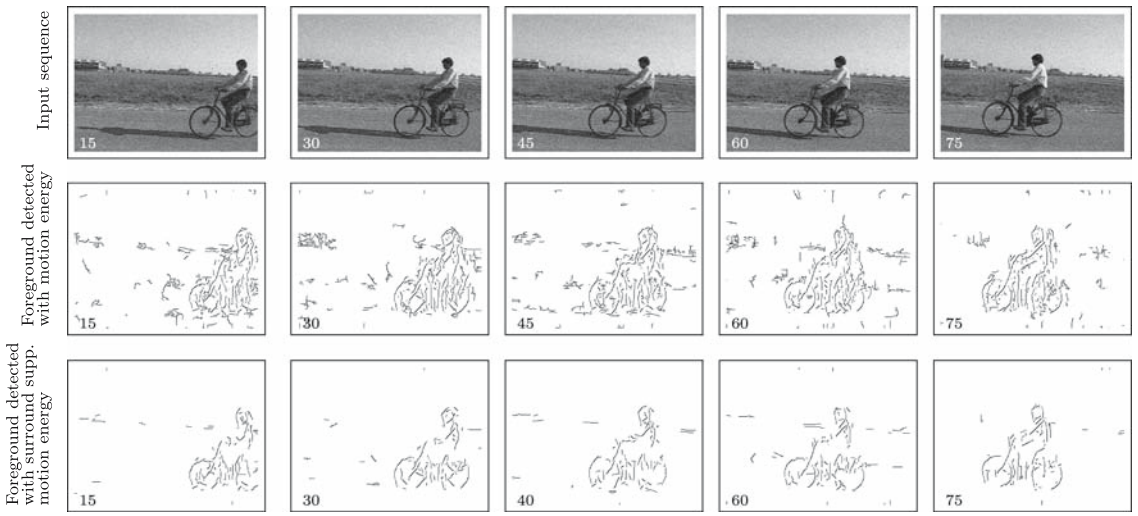

Fig. 12 Motion detection. The *top row* shows a video sequence which consists of a moving object in a stationary background (signal-to-noise ratio, SNR = 20 dB). Subsequent rows show the binary contours of the moving object, which are computed by determining whether the response of a motion energy filter (*second row*), and a surround suppressed motion energy filter (computed with $\alpha = 2$, *third row*) with a preferred nonzero speed is higher than the response of the respective filter with a preferred zero speed. Binarization is achieved by non-maxima suppression followed by hysteresis thresholding (Canny 1986), with the most favorable threshold value $t_h = 0.05$ ($t_l = 0.5 \times t_h$)

to moving stimuli and vice versa a filter that prefers a moving stimulus will also respond to a stationary one. This ambiguity of the separate filters regarding the presence or absence of motion in a given position is in contrast with the sharp distinction that our visual system can make between the two conditions. Evidently, the presence or absence of motion at a given position is coded in the set of responses of multiple filters at that position, a situation that is referred to as population coding (Pouget et al. 2000). We show that motion at a given spatial position can be detected in a straightforward way from the motion energy population code responsible for that position. If motion is present, a filter that prefers a stationary stimulus ($v = 0$) will give a smaller response than another filter that is tuned to a preferred nonzero speed. This is demonstrated in Fig. 12 which shows several frames of a video sequence with added random Gaussian noise (SNR = 20 dB) and the edge positions in which a (motion energy and a surround suppressed motion energy) filter with a preferred nonzero speed has a higher response than the identically oriented filter with a preferred zero speed. As can be seen from the figure, this straightforward decoding scheme gives reasonable results for detecting motion. Due to the higher robustness to noise of the surround suppressed motion energy operator the results obtained with that operator contain less false-positive responses in the stationary background.

3.3 Contour detection by surround suppression of texture

In previous works (Petkov and Westenberg 2003; Grigorescu et al. 2003), in which purely spatial 2D Gabor filters were used, it was suggested that surround inhibition facilitates the detection of object contours and region boundaries by suppressing response to texture. Here, we suggest the same biological utility for the spatiotemporal model of cortical cells considered above.

Figure 13 illustrates the effect of surround suppression compared to spatial Gabor energy and motion energy filters. The first row shows frames from a video sequence with added Gaussian noise (SNR = 26 dB) that is uncorrelated from frame to frame. The associated ground truth frames that contain object contours specified by a human observer are displayed in the second row. The purely spatial 2D Gabor energy filter (third row) that is applied on a frame-by-frame basis is not robust to noise. In contrast, the spatiotemporal 3D motion energy filter successfully deals with noise by means of temporal integration that increases the SNR in the output. Yet, the motion energy (fourth row) operator detects all moving edges disregarding their origin: textures, object contours and region boundaries. The spatiotemporal surround suppressed motion energy operator (fifth row) makes a difference between these two types of edges—it inhibits texture edges while preserving object contours and region boundaries.

In this way, the visibility of object contours is increased and this facilitates the detection and recognition of objects. Moreover, surround inhibition also helps eliminating information about uniform motion. Typically, the motion flow generated due to eye or body movement contains mostly trivial information. Such motion information should be suppressed and surround suppression may play an important role in performing this task.

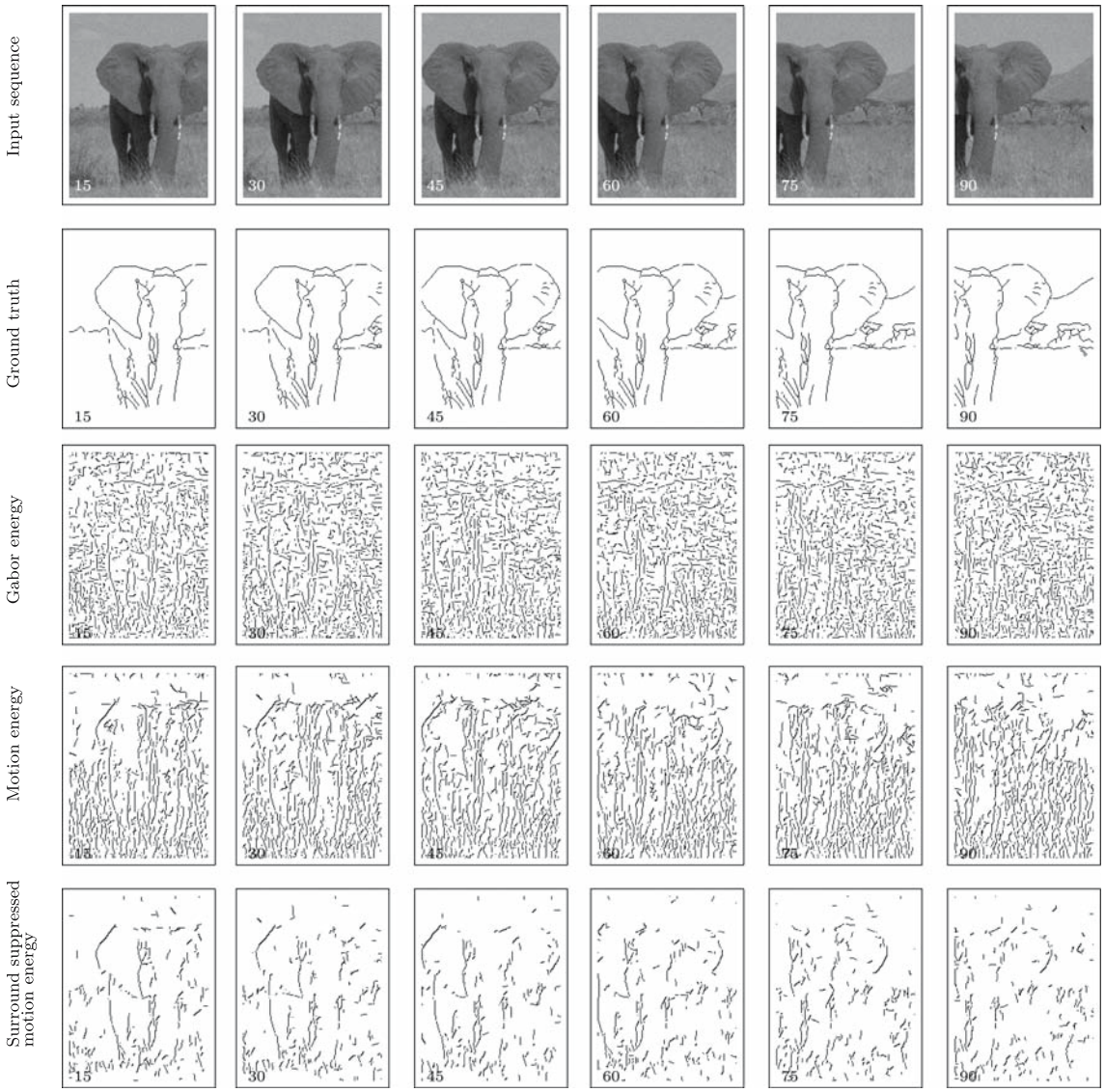

Fig. 13 Edge and contour detection with Gabor-function-based filters. *First row*: frames of an input sequence with uniform motion generated by a frame window sliding over a stationary image (SNR = 26 dB). Ground truth is displayed in the *second row*. Subsequent rows contain the binarized outputs of various operators. *Third row*: Gabor energy ($t_h = 0.08$). *Fourth row*: motion energy ($t_h = 0.045$). *Fifth row*: surround suppressed motion energy ($\alpha = 2$, $t_h = 0.03$)

4 Discussion

In our 3D Gabor function model, a stationary or a moving Gaussian envelope can be used. So far, we have not discussed which of these two options is more appropriate. DeAngelis and co-workers (DeAngelis et al.1993a, b, 1995) mention that their electrophysiological results suggest a stationary envelope. Van Hateren and Ruderman (van Hateren and Ruderman 1998) make a similar suggestion based on independent component analysis of video sequences of natural scenes. In both cases, however, there is no quantitative analysis of this aspect of spatiotemporal receptive fields and the suggestions seems to have been based on a qualitative visual inspection of the obtained profiles. Furthermore, the quantitative results of (DeAngelis et al. 1993b) for direction tuning point to a moving rather than stationary envelope if considered in the context of the respective orientation tuning curves (Fig. 6a). Our computer simulations show that this question is of secondary importance for the functional aspects studied in this work and that all results reported above hold qualitatively for both a stationary and a moving envelope. The speed of the Gaussian envelope only affects the strength of the response for optimal versus nonoptimal stimulus direction and speed, but it has no influence on the preferred speed, which

is determined only by the phase speed of the moving cosine wave. The origin of direction and speed tuning properties, although not addressed in the current models, can be due to linear superposition of geniculate and intracortical contributions (Sabatini and Solari 1999). Further, these models have a functional link to the classical Reichardt model (Reichardt 1961) because of their relation to the energy model (Adelson and Bergen 1985).

In Sect. 3.2 we show that the population code generated by a set of 3D Gabor filters tuned to different preferred speeds can be used in a straightforward way to detect motion. The sign of the difference between the response of a motion energy filter with a preferred nonzero speed and an identically oriented filter that prefers stationary stimuli is indicative of the presence or absence of motion in a given position. However, we are not aware of any neural correlate of such a computation.

The spatiotemporal filters discussed in the current work are inspired by the properties of V1 cells. Typically, V1 cells have small receptive fields and therefore can see only the component of motion that is orthogonal to the orientation of a moving edge; this is known as the aperture problem (Movshon et al. 1985; Heeger 1987). There are several theories which speculate on how and where pattern motion is computed from V1 outputs. One idea is that pattern motion is computed in MT (Adelson and Movshon 1982; Albright 1984; Heeger 1987; Movshon et al. 1985; Simoncelli and Heeger 1998) where the V1 outputs are combined using the intersection of constraints (IOC) rule (Adelson and Movshon 1982; Simoncelli and Heeger 1998) or vector averaging (Mingolla et al. 1992; Rubin and Hochstein 1993). Another idea is that end-stopped cells in V1 could be involved in encoding pattern motion because they respond well to line terminators (or features) moving in their preferred direction and speed, independent of the orientation of the contour (Pack and Born 2001; Pack et al. 2003; Born and Bradley 2005). In this case, MT cells just need to combine V1 outputs. In addition, network models incorporated with feedback mechanisms have also been proposed to support the idea that pattern motion can be computed at the V1 stage itself (Bayerl and Neumann 2004; Bayerl and Neumann 2007).

The proposed model for surround suppression possesses similarities with certain mechanisms that were suggested in the literature for a different purpose. For instance, in a model of simple cells proposed by (Heeger 1993), there is a normalization stage in which the response of a cell is divided by the pooled activity of a large number of cells. This divisive normalization mechanism successfully accounted for response saturation for high-contrast stimuli exhibited by many cortical cells (Tolhurst and Dean 1991) and for certain aspects of direction tuning. Our surround inhibition scheme is related to the normalization model in the sense that the activity of a cell is suppressed by the responses of other cells in a certain neighborhood. However, the main difference lies in our motivation, which is to explore other functional consequences of surround mechanisms in the spatiotemporal domain, viz. noise reduction, texture suppression, improved contour visibility, and figure/ground segregation. On the modeling side, there is a difference concerning the inclusion or exclusion of the area of the CRF in computing the inhibition term. This is a model design issue since the existing electrophysiological studies (Knierim and van Essen 1992; Nothdurft et al. 1999) exclude the CRF region from suppression measurements and therefore cannot conclusively answer the question of whether or not suppression originates from the CRF. The results of some anatomic studies on the distribution of horizontal interconnections in area V1 show that the sites (buttons) at which a neuron connects to other neurons are located outside a certain area around the considered neuron (Bosking et al. 1997). This may point to exclusion of the CRF area from the surround weighting function. In Fig. 14 we demonstrate how the results would change if a surround weighting function that covered the CRF and its surroundings were used. In areas with texture the inhibition term is similar for both models. For contours (as represented by an isolated bar), however, the inclusion of the CRF area in the support of the surround weighting function leads to a higher self-inhibition. This self-inhibition is reduced by exclusion of the CRF area from the surround weighting function in our model. It can be further reduced by exclusion of further areas from the CRF surround that are collinear with the optimal center stimulus

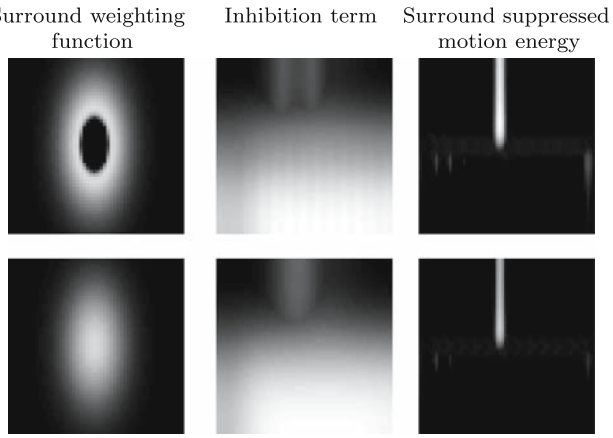

Fig. 14 Each row shows a snapshot of a surround weighting function (*left*), an inhibition term computed with this function for the stimulus shown in Fig. 9 (*middle*), and the output of the corresponding surround suppressed motion energy operator (*right*). The *top row* shows the surround suppression calculated using a surround weighting function that excludes the CRF while the *bottom row* shows the surround inhibition calculated using a weighting function that includes the CRF. In areas with texture the inhibition term is similar for both models. However, for contours (as represented by the isolated bar) there is higher self-inhibition if the CRF area is included in the inhibition (*bottom row*). Consequently, the responses to contours are smaller. However, qualitatively the results are similar

(Papari et al. 2007). Actually, neurophysiological studies (Xiao et al. 1995, 1997a, b, 1998) suggest that about half of the antagonistic surrounds in MT/V5 are asymmetric, with most of the suppression being confined to a single side of the receptive field (Born and Bradley 2005). In any case, noise and texture will be suppressed stronger than contours.

A point that deserves special attention for clarifying the properties of our model is the setting of the surround suppression parameter α in Eq. 6. In all illustrations presented in this paper we used the value $\alpha = 2$. There is a theoretical reason for this choice, related to the fact that the proposed inhibition scheme (Eq. 6) is a subtractive linear mechanism followed by a nonlinear half-wave rectification. Consider the periodic grating in the lower part of the input sequence shown in Fig. 9. The response of the motion energy operator is also a grating-like structure with alternating crests and troughs. Since the weights of the inhibition kernel are normalized using the L_1 norm to give an integral of 1, a value of $\alpha \geq 2$ is necessary to compute an inhibition term that can completely suppress the crests in the motion energy response.[2] In practice, the value of $\alpha = 2$ is sufficient to suppress the periodic grating structure present in the input. As illustrated in Fig. 11, this value of α is appropriate for eliminating noise as well. Higher values of α are not desirable because they lead to increased suppression of object contours. Hence, the appropriate choice of α is a balance of contradicting design issues: suppression of noise and texture (favored by high values of α) versus retainment of object contours (favored by low values of α). One can use the metric defined in Eq. 7 to arrive at a particular choice of α. Let DC be the set of points identified as being part of the contour by a given contour detector (see rows three, four and five in Fig. 13) and GT be the set of contour pixels in the corresponding ground truth image (see second row in Fig. 13). We define recall (R) and precision (P) as follows:

$$R = \frac{\text{card}\{DC \cap GT\}}{\text{card}\{GT\}}, \qquad P = \frac{\text{card}\{DC \cap GT\}}{\text{card}\{DC\}} \tag{7}$$

where card$\{X\}$ is the number of elements in set X and the intersection of GT and DC is computed to compensate for small shifts of contours detected by an operator (Grigorescu et al. 2003; Papari et al. 2007). Only those values of α that produce reasonably large values of both recall and precision are interesting. This is illustrated in Fig. 15, where recall and precision values are shown for different values of the suppression parameter α for the sequence shown in Fig. 13. For each value of α, the binarized output was computed using

[2] A homogeneous response field would be adequately suppressed for $\alpha \geq 1$. The response to texture is, however, never a homogeneous field but rather shows a crest–trough structure.

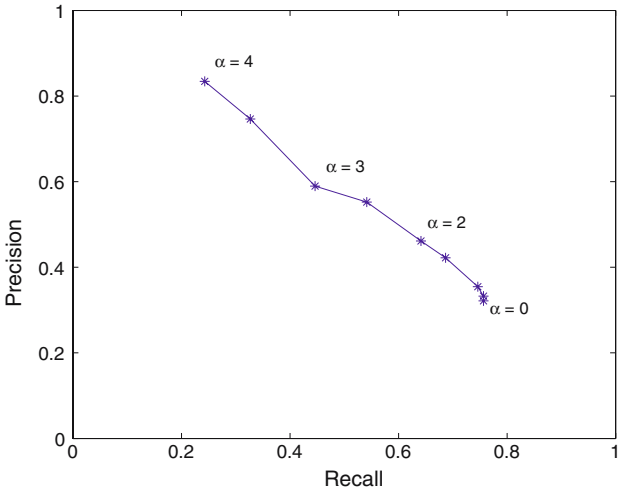

Fig. 15 On the choice of α: plot of precision and recall values for different values of α for the elephant sequence shown in Fig. 13. Low values of α yield high recall and low precision, and the situation is reversed for high values of α. Intermediate values of α lying between 2 and 3 produces reasonable recall and precision values. The harmonic mean of precision and recall reaches its maximum for $\alpha = 2.5$

suitable threshold value and the values of precision (P) and recall (R) were calculated for each frame and their averages over all frames are used for the plot. One can observe that low values of α yield high recall (i.e., good contour retainment) and low precision (i.e., lots of response to noise and texture) and the situation is reversed for high values of α. Intermediate values of α lying between 2 and 3 produce reasonable recall and precision values. The location of the maximum of the harmonic mean of precision (P) and recall (R) along the curve can be used to identify the optimal parameter value (α) for a given input sequence (van Rijsbergen 1979; Martin et al. 2004).

The simulation results shown in Fig. 9 suggest that the responses to a moving oriented texture pattern will be suppressed. This is due to the fact that the center–surround interactions of our model neurons are antagonistic in nature. We emphasize that our model concerns V1 cells as already pointed out in the introduction. As shown in (Knierim and Van Essen 1992; Nothdurft at al. 1999) the majority of orientation-selective cells in V1 exhibit surround inhibition that leads to suppression of responses of texture. However, there are also neurons in V1 and V2, called grating cells, that show selective responses to oriented texture (von der Heydt et al. 1991, 1992; du Buf 2007,Kruizinga and Petkov 1999). Furthermore, there are cells in MT, called wide-field neurons, that prefer large moving texture fields and exhibit no surround inhibition (Allman et al. 1985; Born and Tootell 1992; Raiguel et al. 1995). Our model is not aimed at reflecting the properties of these types of neurons, nor of neurons in the MT or MST in general. It is believed that wide-field neurons codify background motion and center–surround neurons

specify object motion (Born et al. 2000; Berezovskii and Born 2000). In this context, the results obtained in this work add support to the claim that surround mechanisms help segregate figure from background. Some further experimental/perceptual evidence also exist to support this idea. For instance, it has been reported that surround suppression mechanisms in old people and patients with schizophrenia are weak. At the same time, they experience difficulties in segregating figure from background (Betts et al. 2005; Tadin et al. 2006), a finding that underlies the importance of surround mechanisms.

We also note that the visual system captures information at multiple scales and generally the whole scale space (Koenderink 1984) is used for performing various tasks (ter Haar Romeny 2003; Rodrigues and du Buf 2006). In our scheme, filters that prefer higher speeds have bigger receptive fields and therefore the motion detection mechanism proposed in Sect. 3.2 has a multiscale aspect. Elsewhere, the concept of surround inhibition has been used in a multiscale approach for enhancing contour detection in purely spatial images (Papari et al. 2006).

5 Conclusions

Spatiotemporal (3D) Gabor filters applied to video sequences have advantages over purely spatial (2D) Gabor filters applied on a frame-by-frame basis.

First, spatiotemporal filters are much more effective at reducing noise compared to purely spatial filters. This is due to the fact that, while noise is uncorrelated from frame to frame, a moving stimulus shifts at a given speed. An appropriate spatiotemporal filter tuned to that speed combines information about the signal from the past frames in a coordinated way to produce the current output frame. Thus, processing using such a filter is more beneficial for the signal than for the noise and this leads to an increased signal-to-noise ratio in the filter output.

Second, motion is an inherently spatiotemporal phenomenon and cannot be dealt with by purely spatial filters on a frame-by-frame basis. Spatiotemporal Gabor filters inspired by the function of simple and complex cells can be used for processing motion. Such filters are broadly tuned to speed: while having some preferred speed, they respond not only to that speed but also to stimuli moving at different velocities as well as stationary stimuli. Therefore, a single such filter does not provide enough information to answer the question of whether there is motion at a given spatial position. That information is population coded in the responses of a group of filters at the position concerned. The presence (or absence) of motion can however be inferred from the population code in a straightforward way. This is so because, at spatial positions where there is movement, a filter with a preferred nonzero speed gives a higher response than an identically oriented filter with a preferred zero speed and a simple comparison of the filter responses suffices to detect motion.

Third, with respect to the biological utility of surround inhibition our results suggest that this mechanism leads to reduced responses to texture while not affecting the responses to object contours and region boundaries. It also further reduces the influence of noise. In this way, the contours of moving objects that are embedded in natural scenes that are rich in texture become more visible, which facilitates the detection and recognition of objects in such scenes and segregation of figure from their backgrounds. Another important biological utility that surround suppression might have is to suppress uniform motion generated by the background due to head or eye movement.

We believe that, although the current model is based on motion sensitive neurons in V1, it provides a general framework to model surround interactions at all levels. Next to improving the understanding of motion processing in the visual system of man and animals, the insights gained from the computational models proposed above can be used in computer vision algorithms.

Appendix A: Mathematical details of the CRF function

We define the CRF function of a model simple cell, $g_{v,\theta,\varphi}(x, y, t)$, $(x, y, t) \in \Omega \subset \mathbb{R}^3$, which is centered in the origin $(0, 0, 0)$ as follows:

$$
\begin{aligned}
g_{v,\theta,\varphi}(x, y, t) = \frac{\gamma}{2\pi\sigma^2} \exp \frac{-\left((\bar{x} + v_c t)^2 + \gamma^2 \bar{y}^2\right)}{2\sigma^2} \\
\times \cos\left(\frac{2\pi}{\lambda}(\bar{x} + vt) + \varphi\right) \\
\times \frac{1}{\sqrt{2\pi}\tau} \exp -\frac{(t - \mu_t)^2}{2\tau^2} U(t) \\
\bar{x} = x \cos(\theta) + y \sin(\theta) \\
\bar{y} = -x \sin(\theta) + y \cos(\theta) \\
U(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0, \end{cases}
\end{aligned} \tag{8}
$$

where the parameter γ is the spatial aspect ratio that specifies the ellipticity of the Gaussian envelope factor in the spatial domain. The standard deviation σ of this Gaussian factor determines the size of the receptive field. The parameter v_c is the speed with which the center of the spatial Gaussian envelope moves along the $\bar{x}$ axis. When $v_c = 0$, the center of the Gaussian envelope is stationary. The parameter λ is the spatial period or wavelength and $1/\lambda$ the spatial frequency of the cosine factor. The angle parameter $\theta \in [0, 2\pi)$ determines the preferred direction of motion and the preferred spatial orientation of the filter. For instance,

when $\theta = 0$, a vertical edge moving rightwards will evoke a higher response than edges of other orientations and directions of movement. The parameter v is the phase speed of the cosine factor and determines the preferred speed of motion. The phase offset $\varphi \in (-\pi, \pi]$ determines the symmetry of $g_{v,\theta,\varphi}(x, y, t)$ in the spatial domain with respect to its moving center $(\bar{x} + v_c t, \bar{y})$. It is symmetric when $\varphi = 0$ and $\varphi = \pi$ and antisymmetric when $\varphi = -\pi/2$ and $\varphi = \pi/2$. Other values of φ correspond to asymmetric mixtures. We use another Gaussian distribution, with a mean μ_t and standard deviation τ, to model the change in intensities of the excitatory and inhibitory lobes of the receptive field with time. Finally, the unit step function $U(t)$ ensures that the filter is causal and hence considers inputs only from the past.

We now specify the choice of parameter values that is used in the current work. The parameterization that we use to model the spatial properties follows previous works (Petkov and Kruizinga 1997; Kruizinga and Petkov 1999; Petkov and Westenberg 2003; Grigorescu et al. 2003) and takes into account some restrictions found in experimental data. The spatial aspect ratio is set to $\gamma = 0.5$ for which the support of the receptive field is elongated along the $\bar{y}$ axis. The ratio σ/λ determines the spatial bandwidth and the number of excitatory and inhibitory stripe zones in the receptive field. The half-response spatial frequency bandwidth b (in octaves) and the ratio σ/λ are related as follows:

$$\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^b + 1}{2^b - 1}. \tag{9}$$

In this paper, we fix the value of the ratio $\sigma/\lambda = 0.56$, which corresponds to a half-response bandwidth of one octave. We set $v_c = 0$ or $v_c = v$ to obtain a filter with a stationary or a moving envelope, respectively. We use the following relation between the preferred spatial wavelength λ and the preferred speed v: $\lambda = \lambda_0\sqrt{1 + v^2}$, where the constant λ_0 is the spatiotemporal period of the filter. In this work, we choose $\lambda_0 = 2$ which is the minimum spatiotemporal period that could be used in digital image sequences. The above relation between λ and v ensures that we have a family of receptive field functions with a constant spatiotemporal period λ_0. The relation also implies that filters that prefer high speeds have bigger receptive fields. Assuming that image sequences are sampled at a video rate of 25 Hz and one time unit corresponds to 40 ms, we choose $\mu_t = 1.75$ to reflect the fact that the mean time delay of the peak of the receptive field is about 70 ms after the stimulus onset (DeAngelis et al. 1993a). We set $\tau = 2.75$, which corresponds to the observation that the mean duration of most RFs of the concerned type is about 300 ms (DeAngelis et al. 1993a).

The parameters v and θ specify the preferred speed and the direction selectivity of the filter. At the same time, v determines the preferred wavelength (via the relation $\lambda = \lambda_0\sqrt{1 + v^2}$ and the receptive field size (via the relation $\sigma = 0.56\lambda$). Similarly, θ specifies the preferred spatial orientation of the filter.

For a multichannel application like the one described in Sect. 3.2 where responses of filters with different preferred speeds are compared one should in principle carry out an additional normalization of the function $g_{v,\theta,\varphi}(x, y, t)$ such that the filter gives a fixed response to a corresponding optimal stimulus like a step edge moving at speed v in direction θ. Such a normalization would however not qualitatively change the results displayed in Figs. 6 and 12.

Appendix B: Mathematical details of the surround suppression weighting function

The surround suppression weighting function is defined as follows:

$$w_{v,\theta,k_1,k_2}(x, y, t) = \frac{I_{v,\theta,k_1,k_2}(x, y, t)}{\|I_{v,\theta,k_1,k_2}\|_1}, \tag{10}$$

where $\|.\|_1$ denotes the L_1 norm and the term $I_{v,\theta,k_1,k_2}(x, y, t)$ is defined as follows:

$$I_{v,\theta,k_1,k_2}(x, y, t) = |G_{v,\theta,k_2}(x, y, t) - G_{v,\theta,k_1}(x, y, t)|^+,$$

$$G_{v,\theta,k}(x, y, t) = \frac{\gamma}{2\pi(k\sigma)^2}\exp\frac{-\left((\bar{x} + v_c t)^2 + \gamma^2\bar{y}^2\right)}{2(k\sigma)^2} \times \frac{1}{\sqrt{2\pi}\tau}\exp-\frac{(t - \mu_t)^2}{2\tau^2}U(t). \tag{11}$$

Observe that the term $G_{v,\theta,k}(x, y, t)$ is similar to the receptive field function $g_{v,\theta,\varphi}(x, y, t)$ but without the cosine factor. The parameters $(\sigma, \gamma, \mu_t, \tau, v_c, \theta)$ have the same functional role and are fixed in the same way as outlined in Appendix A. In this paper, we set $k_1 = 1$ and $k_2 = 4$ and denote the resulting function in (10) as $w_{v,\theta}(x, y, t)$ in the main text.

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2(2):284–299

Adelson EH, Movshon JA (1982) Phenomenal coherence of moving visual patterns. Nature 300(5892):523–525

Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 52(6):1106–1130

Allman JM, Miezin FM, McGuinness E (1985) Direction and velocity specific responses from beyond the classical receptive field in the middle temporal visual area (MT). Perception 14(2):105–126

Andrews BW, Pollen DA (1979) Relationship between spatial frequency selectivity and receptive field profile of simple cells. J Physiol Lond 287:163–176

Bayerl P, Neumann H (2004) Disambiguating visual motion through contextual feedback modulation. Neural Comput 16:2041–2066

Bayerl P, Neumann H (2007) A fast biologically inspired algorithm for recurrent motion estimation. IEEE Trans Pattern Anal Mach Intell 29(2):246–260

Berezovskii VK., Born RT (2000) Specificity of projections from wide field and local motion-processing regions with the middle temporal visual area of the owl monkey. J Neurosci 20(3):1157–1169
Betts LR, Taylor CP, Sekuler AB, Bennett PJ (2005) Aging reduces center-surround antagonism in visual motion processing. Neuron 45:361–366
Blakemore C, Tobin EA (1972) Lateral inhibition between orientation detectors in the cat's visual cortex. Exp Brain Res 15:439–440
Born RT (2000) Center–surround interactions in middle temporal visual area of the owl monkey. J Neurophysiol 84(5):2658–2669
Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189
Born RT, Tootell RBH (1992) Segregation of global and local motion processing in primate middle temporal visual area. Nature 357(6378):497–499
Born RT, Groh JM, Zhao R, Lukasewycz SJ (2000) Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. Neuron 26:725–734
Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew stirate cortex. J Neurosci 17(6):2112–2127
Bradley DC, Anderson RA (1998) Center-surround antagonism based on disparity in primate area MT. J Neurosci 18(18):7552–7565
Buracas GT, Albright TD (1996) Contribution of area MT to perception of three dimensional shape: a computational study. Vis Res 36(6):869–887
Canny JF (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8(6):679–698
Daugman JG (1985) Uncertainty relations for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters. J Opt Soc Am A 2:1160–1169
DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. J Neurophysiol 89(2):1094–1111
DeAngelis GC, Ohzawa I, Freeman RD (1993a) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. J Neurophysiol 69:1091–1117
DeAngelis GC, Ohzawa I, Freeman RD (1993b) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. J Neurophysiol 69:1118–1135
DeAngelis GC, Ohzawa I, Freeman RD (1995) Receptive-field dynamics in the central visual pathways. Trends Neurosci 18:451–458
du Buf JMH (2007) Improved grating and bar cell models in cortical area V1 and texture coding. Image Vis Comput 25(6):873–882
Eifuku S, Wurtz RH (1998) Response to motion in extrastriate cortex MSTl: center–surround interactions. J Neurophysiol 80(1):282–296
Gautama T, van Hulle MM (2001) Function of center–surround antagonism for motion in visual area MT/V5: a modeling study. Vis Res 41(28):3917–3930
Glezer VD, Tscherbach TA, Gauselman VE, Bondarko VM (1980) Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex. Biol Cybern 37:195–208
Grigorescu C, Petkov N, Westenberg MA (2003) Contour detection based on nonclassical receptive field inhibition. IEEE Trans Image Process 12(7):729–739
Grigorescu C, Petkov N, Westenberg MA (2004) Contour and boundary detection improved by surround suppression of texture edges. Image Vis Comput 22(8):609–622
Grigorescu SE, Petkov N, Kruizinga P (2002) Comparison of texture features based on Gabor filters. IEEE Trans Image Process 11(10):1160–1167
Heeger DJ (1987) Model for the extraction of image flow. J Opt Soc Am A 4(8):1455–1471
Heeger DJ (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. J Neurophysiol 70(5):1885–1898
Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. J Physiol Lond 160:106–154
Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195(1):215–243
Jones HE, Grieve KL, Wang W, Silito AM (2001) Surround suppression in primate V1. J Neurophysiol 86:2011–2028
Jones JP, Palmer LA (1987) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat stirate cortex. J Neurophysiol 58:1233–1258
Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67:961–980
Koenderink JJ (1984) The structure of images. Biol Cybern 50:363–370
Koenderink JJ, van Doorn AJ (1992) Second order optic flow. J Opt Soc Am A 9:530–538
Kruizinga P, Petkov N (1999) Nonlinear operator for oriented texture. IEEE Trans Image Process 8(10):1395–1407
Kulikowski JJ, Bishop PO (1981) Fourier analysis and spatial representation in the visual cortex. Experientia 37:160–163
Mahalanobis PC (1936) On the generalised distance in statistics. Proc Natl Inst Sci India 12:49–55
Marcelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70(11):1297–1300
Martin DR, Fowlkes CC, Malik J (2004) Learning to detect natural image boundaries using local brightness, color and texture cues. IEEE Trans Pattern Anal Mach Intell 26(5):530–549
Mingolla E, Todd JT, Norman JF (1992) The perception of globally coherent motion. Vis Res 32(6):1015–1031
Morrone MC, Burr DC (1988) Feature detection in human vision: a phase-dependent energy model. Proc R Soc Lond B 235:221–245
Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. Experimental brain research supplemntum II: pattern recognition mechanisms pp 117–151
Movshon JA, Thompson ID, Tolhurst DJ (1978a) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99
Movshon JA, Thompson ID, Tolhurst DJ (1978b) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol 283:53–77
Nakayama K, Loomis JM (1974) Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. Perception 3:63–80
Nothdurft HC, Gallant JL, van Essen DC (1999) Response modulation by texture surround in primate area V1: correlates of "popout" under anasthesia. Vis Neurosci 16:15–34
Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature 409(6823):1040–1042
Pack CC, Hunter JN, Born RT (2005) Contrast dependence of suppresive influences in cortical area MT of alert macaque. J Neurophysiol 93(3):1809–1815
Pack CC, Livingstone MS, Duffy KR, Born RT (2003) End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. Neuron 39(4):671–680
Paffen CLE, van der Smagt MJ, te Pas SF, Verstraten FAJ (2005) Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. J Vis 5:571–578
Papari G, Campisi P, Petkov N, Neri A (2006) A multiscale approach to contour detection by texture suppression. In: Dougherty ER,

Astola JT, Egiazarian KO, Nasrabadi NM, Rizvi SA (eds) Image processing: algorithms and systems, neural network, and machine learning; Proceedings SPIE-IST Electronic Imaging 2006, San Jose, CA, USA, January 16–18, 2006, vol 6064, SPIE, Bellingham, Washington; IST, Springfield, Virginia, pp 60640D-1– 60640D-12

Papari G, Campisi P, Petkov N, Neri A (2007) A biologically motivated multiresolution approach to contour detection. EURASIP J Adv Signal Process 2007 (Article ID 71828), 28 pp

Petkov N, Kruizinga P (1997) Computational models of visual neurons specialized in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. Biol Cybern 76(2):83–96

Petkov N, Westenberg MA (2003) Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition. Biol Cybern 88(10):236–246

Pouget A, Zemel RS, Dayan P (2000) Information processing with population codes. Nat Rev Neurosci 1(2):125–132

Raiguel SE, van Hulle MM, Xiao DK, Marcar VL, Orban GA (1995) Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. Eur J Neurosci 7(10):2064–2082

Reichardt W (1961) Autocorrelation, a priniciple for the evaluation of sensory information by the central nervous system. In: Rosenblith WA (ed) Sensory communication. Wiley, New York

Rodrigues J, du Buf JMH (2005a) Multi-scale cortical keypoint representation for attention and object detection, Pattern recognition and image analysis. Proc Lect Notes Comput Sci 3523:255–262

Rodrigues J, du Buf JMH (2005b) Multi-scale keypoints in V1 and face detection Brain, vision and artificial intelligence. Proc Lect Notes Comput Sci 3704:205–214

Rodrigues J, du Buf JMH (2006) Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. Biosystems 86:75–90

Rubin N, Hochstein S (1993) Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. Vis Res 33(10):1385–1396

Sabatini SP, Solari F (1999) An architectural hypothesis for direction selectivity in the visual cortex: the role of spatially asymmetric intracortical inhibition. Biol Cybern 80(3):171–183

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vis Res 38(5):743–761

Spitzer H, Hochstein S (1985) A complex cell receptive field model. J Neurophysiol 53:1266–1286

Tadin D, Kim J, Doop ML, Gibson C, Blake R, Lappin JS, Park S (2006) Weakened center-surround interactions in visual motion processing in schizophrenia. J Neurosci 26:11403–11412

Tadin D, Lappin JS, Gilroy LA, Blake R (2003) Perceptual consequences of center–surround antagonism in visual motion processing. Nature 424(6946):312–315

Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, Iwai E (1986) Analysis of local and widefield movements in the superior temporal visual areas of the macaque monkey. J Neurosci 6(1):134–144

ter Haar Romeny BM (2003) Front-end vision and multi-scale image analysis: multi-scale computer vision theory and applications, written in mathematica. Kluwer, Dordrecht

Tolhurst DJ, Dean AF (1991) Evaluation of a linear model of directional selectivity in simple cells of the cat's stirate cortex. Vis Neurosci 6:421–428

van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B 265:2315–2320

van Rijsbergen C (1979) Information retrieval, 2nd edn. Department of Computer Science, University of Glasgow

von der Heydt R, Peterhans E, Dursteler MR (1991) Grating cells in monkey visual cortex: coding texture. In: Blum B (ed) Channels in the visual nervous system: neurophysiology, psychophysics and models. Freund, London, pp 53–73

von der Heydt R, Peterhans E, Dursteler MR (1992) Periodic pattern selective cells in monkey visual cortex. J Neurosci 12:1416–1434

Xiao DK, Marcar V, Raiguel SE, Orban GA (1997a) Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. Eur J Neurosci 9(5):956–964

Xiao DK, Raiguel SE, Marcar V, Orban GA (1998) Influence of stimulus speed upon the antagonistic surrounds of area MT/V5 neurons. NeuroReport 9:1321–1326

Xiao DK, Raiguel SE, Marcar V, Koenderink JJ, Orban GA (1995) Spatial heterogenity of inhibitory surrounds in the middle temporal visual area. Proc Natl Acad Sci USA 92(24):11303–11306

Xiao DK, Raiguel SE, Markar V, Orban GA (1997b) The spatial distribution of the antagonistic surround of MT/V5 neurons. Cereb Cortex 7(7):662–667